A simple crime hotspot forecasting algorithm

Anonymous Author(s)
Affiliation
Address
email

Abstract

Crime hotspot forecasting is an important part of crime prevention and reducing the delay between a 911 call and the physical intervention. Current developments in the field focus on enriching the historical data and sophisticated point process analysis methods with a fixed grid. In the paper we present a simple spatio-temporal point process allowing one to perform exhaustive (literal) grid searches. We then show that this approach can compete with more complex methods, as evidenced by the results on data collected by the Portland Bureau of Police. Finally, we discuss the advantages and potential implications of the new method.

1 Introduction

Spatio-temporal crime forecasting is a field that grabs the attention of both scientists and practitioners. Many academic researchers have published results based on time series analysis ([7]), regression methods ([4], [10], [23]), kernel density estimation ([2], [3], [5], [8], [19], [1]) or self-exciting point processes ([11], [22], [21], [15], [13], [16], [12]). Moreover, the US Government appreciates the impact predictive policing has on society (see [18]).

In a typical crime prediction task, the forecast area is fixed and divided into small sub-regions, called cells. The cells are then scored separately over a given future time window. The ones with the highest rate are chosen as the most dangerous areas and called hotspots. In this article we present a point of view for hotspot forecasting that differs from those which can be found in the literature. We emphasise the simplicity and efficiency of our algorithm for a fixed grid to get an opportunity to check as many grids as possible. We place those attributes over sophisticated methods, with state-of-the-art results in practice. Our models would have won eight categories of the Real-Time Crime Forecasting Challenge conducted by the National Institute of Justice ([17]).

The rest of the paper is organized as follows. In Section 2 we explain our approach in detail. Section 3 contains a comprehensive description of case study of our method on data from the Real-Time Crime Forecasting Challenge. Further comments and summary are placed in Section 4.

2 The model

The choice of grid There is a vast literature available about crime forecasting for a given grid of cells based on past crimes committed (see references in the Introduction). In such a setup, more or less sophisticated methods are applied to predict which fixed parts of the investigated region will experience the highest future rate of crime. Clearly, changing the grid changes the entire task as well and may lead to completely different predictions with different levels of effectiveness in the real world. The choice of grid is really important. However, as far as we know, whenever the cell division is not imposed in advance, searching for a good grid is in practice reduced to *grid search*, *random search* (see [20]) or another primitive method of walking among parametrizations of possible tessellations. The reason there is a lack of 'smarter' grid choosing techniques may be that spatial

Submitted to 32nd Conference on Neural Information Processing Systems (NIPS 2018). Do not distribute.

distributions of crimes committed in urban areas are 'weird': they contain atoms with very high crime rates (related to, for example, large-area stores or shelters for the homeless). Therefore, using the same data-driven algorithm for even very similar grids can cause a huge discrepancy in the qualities of the predictions obtained.

Hence, grid optimization cannot be neglected. However, a good grid parametrization should take into account horizontal and vertical shifts, cell height, width, rotations, shape distortions, grids based on triangles, hexagons, aperiodic tessellations… Taking into consideration the massive number of grids worth checking we concluded there was a need for a very fast but still well performing supervised model for a fixed grid, one that would simply execute a random search on a rich space of grid parameterizations to find the 'optimal' grid. This would yield a better final result than a more sophisticated, but slower algorithm applied to a random set of grids that would be too small to contain any decent tessellation.

Fast algorithm for a given grid The main idea behind our algorithm for a fixed grid is simple: count the past crimes in every cell and mark the cells with 'the worst past' as hotspots. In other words, we assume that if many crimes occurred somewhere, more are likely to happen. This principle may strike some as naive and outdated, but we believe that it is both accurate enough and fast. Up-to-date crime registries are freely available for several US cities. They form the main dataset in data-driven crime forecasting algorithms. One can search for any external data which could affect future crimes, but have not left a trace on those crimes that have already been committed. We are aware that weather, demographics and even social media information (see [23]) are sometimes used in similar contexts. Unfortunately, they significantly increase the model's complexity, often without a guarantee of noticeably improved accuracy. Keeping computations as simple as possible, by using merely historical crime data, enables us to spend more time on selecting the right grid.

We refine the raw algorithm by taking care of data aging and seasonality. Namely, we assign weights to all the past crimes and then sum up the weights of all the crimes in consecutive cells to find the hotspots. The weight of an event decreases exponentially as a function of age (in days) of a crime. The intensity of the decrease is a hyperparameter, tuned with the use of available data to obtain the best results. Also, we boost the weights of crimes committed on the same days of the year as those in the forecasted time span. The power of boosting is a hyperparameter as well.

Moreover, we introduce a primitive 'spatial radiation' of past crimes. For each data point, we put eight of its copies with reduced weights in the corners and in the center of the sides of the rhombus around it. In this way, a 'part' of an event that has occured close to the cell border could fall into a neighboring cell. We chose to use a rhombus because it reflects the Manhattan metric, a reasonable match for North-South-oriented axis grid street plans, of which there are many in US cities. In our opinion, this 'degenerated spatial decay' technique is pretty fast and good enough for working with aggregations of crimes to regular convex cells. The size of the rhombus and reduction of weights of added copies are hyperparameters.

The strict mathematical description of the presented approach, expressed in the language of spatio-temporal point processes (cf., e.g., [15] and references therein), is placed in the supplemental material.

Validation In classic crime forecasting, the score functions taken from the binary classification – ROC/AUC, sensitivity, etc. – are used (see [3]). There are also two newer functions on the market: predictive accuracy index (PAI, [3]) and prediction efficiency index (PEI, [9]). Their definitions and the proposed validation routine can be found in the supplementary material. They all have their disadvantages. Binary classification-based functions are inconvenient if the area of the hot-spots to be forecast is a very small fraction of the investigated jurisdiction, which is typical. As for other functions, PAI favors smaller single cell areas while PEI likes as great a single cell area as possible. For this reason it is impossible to maximize both PAI and PEI with the same grid, which casts doubt on the validity of using either of them. Moreover, PEI is bounded by 1 from above whereas the range of PAI is a positive half line, so they are not directly comparable. Here the lack of a simple universal unbiased score function becomes evident. Nevertheless, our approach is metric-agnostic, therefore any reasonable score function can be applied here.

3 Case study

The competition In September 2016, the National Institute of Justice in the US announced the *Real-Time Crime Forecasting Challenge*. The goal was to predict future crimes in Portland, Oregon. Contestants were asked to divide the area under Portland police jurisdiction (an area roughly 15 by 20 miles) into a grid of small cells (i.e., 250 by 250 feet) and indicate the cells that would have the highest future crime rate - hotspots. Several restrictions on the cells' shape and the total volume of hotspots were imposed.

Four different categories of crime were considered separately: all crimes, burglaries, car thefts and street crimes (including assaults, robberies, shots fired). Five future time spans (starting in March 2017) were involved: one week, two weeks, a month, two months and three months. Hence, there were 20 type/time categories. In each of them, the predictions were compared against the actual state of affairs in Portland using both PAI and PEI. Thus, the competition consisted of $4 \cdot 5 \cdot 2 = 40$ separate sub-competitions in total. Only the best submission was awarded in each of them. Three independent tracks of the challenge were run simultaneously: intended for large businesses, small businesses and students, respectively. Each track had the same rules and goals, but separate contestants and winners. We decided to benchmark ourselves against the results from the large business track, as it was expectedly the most competitive one.

Data The NIJ delivered historical data on all the crimes registered in Portland between March 2012 and February 2017. Almost 1,000,000 records were provided in total. Each of them contained the day the crime was committed, coordinates (with accuracy to one foot) and the type of crime committed. There were no data gaps. A very small portion of data was located outside the competition area.

The distribution of data between crime categories was highly imbalanced: burglaries, car thefts and street crimes were only 0.5%, 1%, and 16.5% of records, respectively. One would anticipate a similar distribution reflected in crimes committed between March and May 2017. Thus, we expected a huge discrepancy in the numbers of crimes committed between particular type/time categories during that period. That was true, two extreme cases were: all the crimes between March and May 2017 - 65,000 records, and burglaries in the first week of March 2017 - only 20 events.

Distributions of crimes in all the categories with a big enough number of events had similar characteristics: they consisted of the 'dense' part looking like a sample from a continuous distribution and the 'discrete' part made from atoms. It seems that although the accuracy of the coordinates of crimes committed was in general one foot, police officers tended to 'discretize' some areas like stores or shelters to a single spatial point next to the entrance to the building/area.

Computations The first attempts showed that in each of the 20 type/time categories the PAI metric was maximized by a lot of small hotspots whereas PEI behaved best for a small number of large hotspots. Hence it was clear that we should not attempt to satisfy both metrics simultaneously. Since each metric formed an independent sub-competition, it was better to have a good score for one metric than mediocre results for both. So, for each of the 20 type/time categories we had to decide which metric to focus on in our further work. We did not check the participants' choices or results to properly simulate the competitions' environment. Moreover, our approach was metric-agnostic. Hence, to choose a metric, we just tossed a coin for each of 20 type/time categories.

We were examining four types of regular grids: parallelogram grids, triangular grids with 3 vertices at a point, triangular grids with 6 vertices at a point, and hexagonal grids. They were parameterized by cell height, width, translations, rotations and bending. No shape proved noticeably better than other ones. Hence, we ultimately decided to only use unrotated rectangular grids, parameterized by cell height, width, horizontal and vertical shift. Finally, the number of predicted hotspots was also a hyperparameter.

The model was implemented as Python scripts with NumPy and PyTorch packages used. The computations were performed on the Intel® AI DevCloud infrastructure with Intel® Xeon Scalable Processors, together with optimized distributions of Python and PyTorch.

We optimized the grid and our model hyperparameters for each of 20 type/time categories separately. The learned values of hyperparameters and their interpretations can be found in the supplementary material.

Results In the contest track for large businesses, our predictions would have proved the most accurate in seven categories with the largest numbers of crimes committed during the test periods. We optimized the model for PAI in three of them and for PEI in another four. Moreover, all of those predictions would have remained on the top after comparing results from the competition's three tracks (for large business, small businesses and students). This would have been the best result among all the competitors, while the runner-up would have achieved four across-track wins. The table gathering the results of the competition can be seen in the supplementary materials.

The results allowed us to conclude that for both the PAI and PEI metrics we were able to find grids and hotspots with quality competing with predictions obtained by authors of more complicated methods described in the literature (cf. [14], [6]). Our approach proved especially effective in categories with the biggest number of crimes committed.

Since different competitors submitted different grids, we are unable to compare algorithms for a fixed grid created by particular contestants. Therefore, we cannot judge whether the good performance of our models was an effect of thoroughly scouring potential grids or the power of simplicity of our algorithm for a fixed grid, or perhaps both. At this point we can only claim that our pipeline fulfilled its task.

4 Discussion

The comparative case study on crime data from Portland, OR, shows that our computation time-oriented approach can compete with more sophisticated crime forecasting methods existing in the literature. This result is somewhat surprising. One may conclude that the spatio-temporal distribution of crimes committed is too complicated to be estimated well enough with the use of parametric methods. Or maybe the choice of the proper grid matters much more than it seems. Moreover, we have no reason to claim that the good performance of our algorithm is a one-shot success valid only for Portland since our model contains no part priorly adapted to any particular city. Unfortunately, we did not have the opportunity to compare the quality of crime forecasts done with use of different methods (including our own) for the same fixed grid. Such research would shed more light on this field.

The advantage of our algorithm for cases with thousands or more crimes to forecast is an interesting phenomenon. It can be attributed to two possible factors: a specific spatial distribution of crimes or computational simplicity. As the number of events increases, the crimes tend to be spatially distributed more regularly, but with the growing importance of single-point peaks. As stated above, for most statistical parametric methods it may be intractable to cover a distribution containing both a continuous and a discrete part. Comparing the performance of different models for a fixed grid would bear this out. On the other hand, sophisticated algorithms can paradoxically struggle to find the optimal grid and hotspots when presented with large volumes of training data. A time-consuming training procedure for a fixed grid does not allow one to check a sufficient number of potential grids. This problem may be addressed by more efficient algorithms' implementations and significantly increasing computing resources. Also, adding more constraints on the admissible grid shapes clearly solves the problem, though it also makes it less universal.

Finally, we note that in the perspective of maintaining and updating the crime forecasting system, using only the historical crime data seems to be a good solution. It is hard to find any non-constant external factor which can both influence future crimes and be easier to predict than crimes themselves. Besides, the impact of any hidden important feature is ultimately reflected in the historical data. Moreover, changes in the spatial crime distribution caused by system-driven preventive police activities may be not easy to manage when external data sources are used for forecasting. At the same time, a forecasting system based on merely historical data is able to simply retune to the current crime distribution.

References

[1] M. A. Boni and M. S. Gerber. Automatic optimization of localized kernel density estimation for hotspot policing. In *Proc. 15th IEEE Int. Conf. Machine Learning and Applications (ICMLA)*, pages 32–38, Dec. 2016. doi: 10.1109/ICMLA.2016.0015.

[2] K. J. Bowers, S. D. Johnson, and K. Pease. Prospective hot-spotting: The future of crime mapping? *The British Journal of Criminology*, 44:641–658, 2004. ISSN 0007-0955. doi: 10.1093/bjc/azh036.

[3] S. Chainey, L. Tompson, and S. Uhlig. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21:4–28, 2008. ISSN 0955-1662. doi: 10.1057/palgrave.sj.8350066.

[4] J. Cohen, W. L. Gorr, and A. M. Olligschlaeger. Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis*, 39:105–127, 2007. ISSN 0016-7363. doi: 10.1111/j.1538-4632.2006.00697.x.

[5] M. Fielding and V. Jones. 'disrupting the optimal forager': Predictive risk mapping and domestic burglary reduction in trafford, greater manchester. *International Journal of Police Science & Management*, 14: 30–41, 2012. ISSN 1461-3557. doi: 10.1350/ijps.2012.14.1.260.

[6] S. Flaxman, M. Chirico, P. Pereira, and C. Loeffler. Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the nij "Real-Time Crime Forecasting Challenge". 2018.

[7] W. Gorr, A. Olligschlaeger, and Y. Thompson. Short-term forecasting of crime. *International Journal of Forecasting*, 19:579–594, 2003. ISSN 0169-2070. doi: 10.1016/s0169-2070(03)00092-x.

[8] W. L. Gorr and Y. Lee. Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology*, 31:25–47, 2015. ISSN 0748-4518. doi: 10.1007/s10940-014-9223-8.

[9] J. M. Hunt. *Do Crime Hot Spots Move? Exploring the Effects of the Modifiable Areal Unit Problem and Modifiable Temporal Unit Problem on Crime Hot Spot Stability*. PhD thesis, 2016.

[10] L. W. Kennedy, J. M. Caplan, and E. Piza. Risk clusters, hotspots, and spatial intelligence: Risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology*, 27:339–362, 2011. ISSN 0748-4518. doi: 10.1007/s10940-010-9126-2.

[11] H. Liu and D. E. Brown. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, 19:603–622, 2003. ISSN 0169-2070. doi: 10.1016/s0169-2070(03)00094-3.

[12] C. Loeffler and S. Flaxman. Is gun violence contagious? A spatiotemporal test. *Journal of Quantitative Criminology*, 2017. ISSN 0748-4518. doi: 10.1007/s10940-017-9363-8.

[13] G. Mohler. Marked point process hotspot maps for homicide and gun crime prediction in chicago. *International Journal of Forecasting*, 30:491–497, 2014. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2014.01.004.

[14] G. Mohler and M. D. Porter. Rotational grid, PAI-maximizing crime forecasts. to appear in Statistical Analysis and Data Mining, 2018.

[15] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106:100–108, 2011. ISSN 0162-1459. doi: 10.1198/jasa.2011.ap09546.

[16] G. O. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110:1399–1411, 2015. ISSN 0162-1459. doi: 10.1080/01621459.2015.1077710.

[17] National Institute of Justice. Real-Time Crime Forecasting Challenge. https://www.nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx, 2017. URL `https://www.nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx`.

[18] W. Perry, B. McInnis, C. Price, S. Smith, and J. Hollywood. Predictive policing: The role of crime forecasting in law enforcement operations. Technical report, RAND Corporation, Santa Monica, 2013.

[19] M. D. Porter and B. J. Reich. Evaluating temporally weighted kernel density methods for predicting the next event location in a series. *Annals of GIS*, 18:225–240, 2012. ISSN 1947-5683. doi: 10.1080/19475683.2012.691904.

[20] L. A. Rastrigin. About convergence of random search method in extremal control of multi-parameter systems. *Automation and Remote Control*, 24:1467–1473, 1963.

[21] G. Rosser and T. Cheng. Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. *Applied Spatial Analysis and Policy*, pages 1–21, 2016. ISSN 1874-463X. doi: 10.1007/s12061-016-9198-y.

[22] M. A. Taddy. Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association*, 105:1403–1417, 2010. ISSN 0162-1459. doi: 10.1198/jasa.2010.ap09655.

[23] X. Wang, M. S. Gerber, and D. E. Brown. Automatic crime prediction using events extracted from twitter posts. In *Social Computing Behavioral - Cultural Modeling and Prediction*, pages 231–238. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-29047-3_28.